

# On Subgraph Isomorphism

Sergey Gubin \*

**Abstract**—Article explicitly expresses Subgraph Isomorphism by a polynomial size asymmetric linear system.

**Keywords:** *Subgraph Isomorphism, Linear modeling, Algorithm, Computational Complexity, NP-complete*

## Introduction

In 1988, Yannakakis proved [1] that the Traveling Salesman Problem's (TSP) polytope cannot be expressed by a polynomial size symmetric linear program, where symmetry means that the polytope is an invariant under node relabeling. Because TSP is a NP-complete problem [2], the theorem holds for all NP-complete problems. The question about the size of asymmetric linear models was left open in [1] and it has remained open since.

This article answers that question. We present an explicit polynomial size asymmetric linear model for Subgraph Isomorphism (SubGI). Since SubGI is a NP-complete problem [3], this result is complimentary to the Yannakakis theorem.

The polynomial size asymmetric linear system is built based on an arbitrary but fixed labeling of graphs involved - hence the system's asymmetry. The polynomial size for the system is achieved by immersing the problem in a space of higher dimension, where variables present relabeling possibilities for vertex couples.

We illustrate our method with several examples. Particularly, we explicitly present polynomial size asymmetric linear programs for TSP and for the Satisfiability Problem for conjunctive normal forms (SAT).

## 1 Subgraph Isomorphism

Let  $G$  be a given graph - we will call it an input. Let  $S$  be another given graph - we will call it a pattern. The problem is whether  $G$  contains a subgraph which is isomorphic to  $S$ . For any given couple of graphs  $(G, S)$ , this decision problem is a SubGI instance. Its size can be estimated by the number of vertices in graph  $G$ .

Any graph may be seen as a relation. So, SubGI may be seen as a finite version of the following general problem: whether a given relation possesses a given property. That explains the theoretical and practical importance of SubGI.

SubGI is a NP-complete problem [3]. The Ullmann algo-

rithm [4] is the best known method to solve the problem. Yet it and other known general methods are inefficient. Up to date, the efficient methods were known only for particular types of graph couples  $(G, S)$  [5, 6, 7, and others].

This article describes a reduction of SubGI to a system of linear equations and inequalities. The reduction's computational complexity and the resulting system's size are polynomial over the size of SubGI. For a given couple of graphs  $(G, S)$ , the resulting system has solutions iff input  $G$  contains a subgraph isomorphic to pattern  $S$ .

As well as for graphs, our reduction works for (multi) digraphs with (multi) loops. We will present the reduction for the multi digraph version of SubGI which is, in many cases, more practical. So, input  $G$  and pattern  $S$  are (multi) digraphs with (multi) loops everywhere below in this article.

Because our system contains a polynomial number of linear equations and inequalities with a polynomial number of unknowns, it can be solved in polynomial time by, for example, the Khachiyan ellipsoid algorithm [8, 9].

## 2 Base polytope

Let  $n$  be a natural number. Let the following variables be unknowns:

$$\begin{aligned} x_{ij\mu\nu} &= x_{jiv\mu} : i, j, \mu, \nu = 1, 2, \dots, n \quad i \neq j \quad \mu \neq \nu \\ y_{ii\mu\mu} &= y_{jj\nu\nu} : i, j, \mu, \nu = 1, 2, \dots, n \end{aligned}$$

In the case of  $n = 1$ , variables  $x_{ij\mu\nu}$  are missing indeed.

Let's consider the following linear system:

$$\left\{ \begin{array}{l} x_{ij\mu\nu} = x_{jiv\mu}, \quad x_{ij\mu\nu} \geq 0 \\ \text{- where } i, j, \mu, \nu = 1, 2, \dots, n, \quad i \neq j, \quad \mu \neq \nu \\ \sum_{\mu=1, \mu \neq \nu}^n x_{ij\mu\nu} = y_{jj\nu\nu}, \\ \text{- where } i, j, \nu = 1, 2, \dots, n, \quad i \neq j \\ \sum_{i=1, i \neq j}^n x_{ij\mu\nu} = y_{jj\nu\nu}, \\ \text{- where } j, \mu, \nu = 1, 2, \dots, n, \quad \mu \neq \nu \\ \sum_{\nu=1}^n y_{jj\nu\nu} = 1, \quad y_{jj\nu\nu} \geq 0 \\ \text{- where } j = 1, 2, \dots, n \end{array} \right. \quad (1)$$

\*sgubin@genesyslab.com

The system can be described with the following box matrix of size  $n \times n$ :

$$B = \begin{pmatrix} Y_{1,1} & X_{1,2} & \dots & X_{1,n} \\ X_{2,1} & Y_{2,2} & \ddots & X_{2,n} \\ \vdots & \ddots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \dots & Y_{n,n} \end{pmatrix}_{n \times n}$$

The  $i$ -th diagonal box in box matrix  $B$  is the following diagonal matrix:

$$Y_{ii} = \text{diag}(y_{i,i,1,1}, y_{i,i,2,2}, \dots, y_{i,i,\nu,\nu}, \dots, y_{i,i,n,n})$$

The  $(i, j)$ -th off-diagonal box in box matrix  $B$  is the following matrix:

$$X_{ij} = \begin{pmatrix} 0 & x_{i,j,1,2} & \dots & x_{i,j,1,n} \\ x_{i,j,2,1} & 0 & \ddots & x_{i,j,2,n} \\ \vdots & \ddots & \ddots & \vdots \\ x_{i,j,n,1} & x_{i,j,n,2} & \dots & 0 \end{pmatrix}_{n \times n}$$

System 1 reflects the following relations between elements of box matrix  $B$ :

- (1)  $B$  is a symmetric matrix:  $X_{ij} = X_{ji}^T$ ;
- (2) The total of each column in matrix  $X_{ij}$  does not depend on  $i$  but only on box column  $j$  and on the column in this box column. The total is the appropriate element in diagonal matrix  $Y_{jj}$ ;
- (3) The total over  $i$  of elements  $x_{ij\mu\nu}$  does not depend on  $\mu$  but only on box column  $j$  and on the column  $\nu$  in this box column. This total is the appropriate element in matrix  $Y_{jj}$  - element  $y_{jj\nu\nu}$ ;
- (4) The  $(j, \nu)$ -th columns in off-diagonal boxes  $X_{ij}$  of box matrix  $B$  constitute a doubly stochastic matrix multiplied by element  $y_{jj\nu\nu}$ .
- (5) The total of all elements in matrix  $Y_{jj}$  is equal 1;
- (6) Due to the matrix's symmetry, all of the above is true in the horizontal direction, too;

System 1 always has solutions. The following solution is minimal in the sense of Euclidean norm - we call it a *center*:

$$x_{ij\mu\nu} \equiv \frac{1}{n(n-1)}, \quad y_{jj\nu\nu} \equiv \frac{1}{n}$$

Obviously, the set of all solutions of system 1 is a convex set. Also, because system 1 is a linear system, the set is a polytope. We call this polytope a *base polytope*.

The following solution of system 1 is a vertex of the base polytope: there is one and only one non-zero element in each box  $Y_{ii}$  and  $X_{ij}$ . Obviously, all non-zero elements in the boxes are equal 1 and they are arranged in a grid of elements in matrix  $B$ , one element per box. We call any such solution of system 1 a *solution grid*.

The following lemma shows that all vertices of the base polytope are solution grids.

**Lemma 1** Any solution of system 1 is a convex combination of solution grids.

**Proof** System 1 consists of linear equations and the following inequalities:

$$x_{ij\mu\nu} \geq 0, \quad y_{jj\nu\nu} \geq 0$$

- where all indexes are in their appropriate ranges. The linear equations have solutions - the center, for example. The solutions constitute a linear subspace in the linear space of all  $n^2 \times n^2$  matrices with real elements. Thus, vertices of base polytope are those points in the linear subspace where the number of variables which equal 0 is maximal possible. It so happens that these points are the solution grids, QED.

### 3 Compatibility matrix

Let digraphs  $G$  and  $S$  be the given SubGI instance  $(G, S)$ . Let  $V_G$  and  $V_S$  be vertex sets of the input and pattern appropriately. Obviously,

$$|V_G| \geq |V_S|$$

- the instance would have resolution "NO" otherwise.

Now, let's add  $|V_G| - |V_S|$  isolated vertices to pattern  $S$ . Let's preserve notion  $S$  for the resulting pattern. Let  $n$  be the number of vertices in the input and the pattern after the addition of isolated vertices:

$$n = |V_G| = |V_S|$$

Obviously, the SubGI instance  $(G, S)$  emerging after the addition of isolated vertices has the same resolution as the original instance indeed.

Let's arbitrarily label/enumerate vertices in input  $G$  and pattern  $S$ . Let  $A_G$  and  $A_S$  be the adjacency matrices of the input and pattern appropriate to the labeling. Obviously, SubGI instance  $(G, S)$  has resolution "YES" iff there exists such a relabeling of pattern  $S$  that all elements of matrix  $A_S$  emerging after that relabeling will be less than or equal to the appropriate elements of matrix  $A_G$ . In other words, SubGI instance  $(G, S)$  has resolution "YES" iff the following integral quadratic system has solutions<sup>1</sup>:

$$A_G \geq X A_S X^T \quad (2)$$

- where  $X$  is the unknown permutation matrix of size  $n \times n$ . Permutation matrix  $X$  presents the unknown vertex relabeling of pattern  $S$  after which the existence of an input's subgraph isomorphic to  $S$  has to become self-evident. Obviously, such a relabeling of  $S$  exists iff  $G$  has at least one subgraph isomorphic to  $S$ .

To solve system 2, let's build the following matrix which we call a *compatibility matrix*.

<sup>1</sup>For two matrices  $A = (a_{ij})$  and  $B = (b_{ij})$  of the same size, relation  $A \geq B$  means that

$$\forall i, j \quad (a_{ij} \geq b_{ij}).$$

Let the input and pattern's adjacency matrices be as follows:

$$A_G = (g_{\mu\nu})_{n \times n}, \quad A_S = (s_{ij})_{n \times n}$$

For each couple of pattern's vertices, let's build a *compatibility box*. The compatibility box for vertices with labels  $i$  and  $j$  is the following matrix  $C_{ij} = (e_{ij\mu\nu})_{n \times n}$ :

$$e_{ij\mu\nu} = \begin{cases} 1, & s_{ij} \leq g_{\mu\nu} \wedge s_{ji} \leq g_{\nu\mu} \\ 0, & s_{ij} > g_{\mu\nu} \vee s_{ji} > g_{\nu\mu} \end{cases} \quad (3)$$

Compatibility box  $C_{ij}$  shows all possible re-enumerations for the pattern's vertices  $i$  and  $j$  with disregard to the rest of the pattern's vertices. Obviously, compatibility boxes  $C_{ii}$  are diagonal matrices. And all diagonal elements in compatibility boxes  $C_{ij}$ ,  $i \neq j$ , are equal 0.

The compatibility matrix for SubGI instance  $(G, S)$  is the following box matrix:

$$C = (C_{ij})_{n \times n}$$

The compatibility matrix aggregates all compatibility boxes in accordance with their indexes.

Obviously, integral quadratic system 2 has a solution iff in the compatibility matrix there is a grid of elements, one element per compatibility box, in which all elements are equal 1:

$$\gamma = \{e_{ij\mu\nu} = 1 \mid \mu = \mu(i), \nu = \nu(j)\}$$

- where  $\gamma$  is the grid. Any such grid of elements in compatibility matrix  $C$  we call a *solution grid*<sup>2</sup>, too.

**Lemma 2** SubGI instance  $(G, S)$  has resolution “YES” iff compatibility matrix  $C$  contains a solution grid.

**Proof** Any solution grid defines a vertex relabeling of  $S$  which satisfies system 2, QED.

## 4 Linear model for SubGI

The similarities between compatibility matrix  $C$  and the base polytope  $B$  are obvious. Due to lemmas 1 and 2, we can decide about the existence/absence of solution grids in matrix  $C$  searching matrix  $B$  for solution grids subject to the following constrains:

$$x_{i_0 j_0 \mu_0 \nu_0} = 0, \quad y_{j_0 j_0 \nu_0 \nu_0} = 0 \quad (4)$$

- where indexes are the indexes of all those elements of compatibility matrix  $C$  which are equal 0,

$$e_{i_0 j_0 \mu_0 \nu_0} = e_{j_0 j_0 \nu_0 \nu_0} = 0$$

Then, lemmas 1 and 2 imply the following polynomial size asymmetric linear model for SubGI.

<sup>2</sup>In the next section, we will show that the solution grids from this section and the solution grids from the previous section are the same.

**Theorem** SubGI instance  $(G, S)$  has resolution “YES” iff the aggregated system 1 and 4 has a solution.

**Proof** Any solution of the aggregated system is a convex hull of solution grids. There is a solution grid iff the resolution for instance  $(G, S)$  is “YES”, QED.

System 1, 4 consists of  $O(n^4)$  linear equations and inequalities with  $O(n^4)$  unknowns. The existence/absence of the system's solutions can be detected using the ellipsoid algorithm [8, 9]. Because all coefficients of the system are 0 or 1, the ellipsoid algorithm will solve this system in strongly polynomial time.

Let's notice that constrains 4 explicitly involve the input and pattern's vertex labeling through their adjacency matrices - see definition 3 of the compatibility boxes. Thus, system 1, 4 is an asymmetric linear system. It can be seen that the system's solutions constitute a convex subset of the Birkhoff polytope [10] in  $R^{n^4}$ . Vertex relabeling of digraphs  $G$  and  $S$  will rotate that subset all over the polytope.

## 5 Examples

Let's use our method and resolve the following SubGI instances.

**Vertex vs vertex:** Let input and pattern have just one vertex each:

$$A_G = (g_{1,1})_{1 \times 1}, \quad A_S = (s_{1,1})_{1 \times 1}$$

System 1 for  $n = 1$  looks as follows:

$$y_{1,1,1,1} = 1$$

Constrains 4 for the instance look as follows:

$$y_{1,1,1,1} = \begin{cases} 1, & s_{1,1} \leq g_{1,1} \\ 0, & s_{1,1} > g_{1,1} \end{cases}$$

Thus, the resolution for this SubGI instance is “YES” iff there is no excess of loops in the pattern:

$$s_{1,1} \leq g_{1,1}$$

**Arc vs arc:** Let input and pattern be just arcs:

$$G : 1 \rightarrow 2, \quad S : 1 \leftarrow 2$$

For  $n = 2$ , system 1 looks as follows:

$$\begin{aligned} x_{1,2,1,2} &= y_{2,2,2,2} \\ x_{1,2,2,1} &= y_{2,2,1,1} \\ x_{2,1,1,2} &= y_{1,1,2,2} \\ x_{2,1,2,1} &= y_{1,1,1,1} \\ y_{1,1,1,1} + y_{1,1,2,2} &= 1 \\ y_{2,2,1,1} + y_{2,2,2,2} &= 1 \end{aligned} \quad (5)$$

Constrains 4 for the given input and pattern may be presented as follows:

$$y_{1,1,1,1} = 0, y_{2,2,2,2} = 0$$

The aggregated system has a solution:

$$\begin{aligned} x_{1,2,2,1} &= y_{2,2,1,1} = x_{2,1,1,2} = y_{1,1,2,2} = 1 \\ x_{1,2,1,2} &= y_{2,2,2,2} = x_{2,1,2,1} = y_{1,1,1,1} = 0 \end{aligned}$$

Thus, the resolution for the given SubGI instance is “YES”. The appropriate relabeling of the pattern is transposition (1, 2).

**Arc vs loop:** Let input and pattern be an arc and a loop appropriately:

$$G : 1 \rightarrow 2, S : 1 \rightarrow 1$$

Adding to  $S$  one isolated vertex with index 2 will produce the case of  $n = 2$ . System 1 for the case is system 5, and constrains 4 for the instance may be presented as follows:

$$y_{1,1,1,1} = 0, y_{1,1,2,2} = 0$$

The aggregated system has no solutions:

$$1 = y_{1,1,1,1} + y_{1,1,2,2} = 0$$

Thus, the resolution for the given SubGI instance is “NO”.

**Arc/loop vs loop/arc:** Let input and pattern be the following digraphs:

$$G : 1 \rightarrow 2 \rightarrow 2, S : 1 \rightarrow 1 \rightarrow 2$$

System 1 for the case is system 5, and constrains 4 for the instance may be presented as follows:

$$y_{1,1,1,1} = 0, x_{1,2,2,1} = 0$$

The aggregated system has no solutions:

$$1 = y_{1,1,2,2} = x_{1,2,2,1} = 0$$

Thus, the resolution for the given SubGI instance is “NO”.

**Edge vs arc:** Let input and pattern be the following digraphs:

$$G : 1 \rightarrow 2 \rightarrow 1, S : 1 \rightarrow 2$$

System 1 for the case is system 5, and there are no constrains 4 for the instance. Thus, the aggregated system consists of system 5 alone. The center of its solutions is the following point:

$$\forall i, j, \mu, \nu (x_{ij\mu\nu} = y_{ii\mu\mu} = 1/2)$$

Thus, the resolution for the given SubGI instance is “YES”.

**Cycle vs edge:** Let input and pattern be the following digraphs:

$$G : 1 \rightarrow 2 \rightarrow 3 \rightarrow 1, S : 1 \rightarrow 2 \rightarrow 1$$

Compatibility matrix  $C$  for this SubGI instance looks as follows:

1	0	0	0	0	0	0	0	1	1
0	1	0	0	0	0	0	1	0	1
0	0	1	0	0	0	0	1	1	0
0	0	0	1	0	0	0	0	1	1
0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	1	0	1	1	0
0	1	1	0	0	1	1	0	0	0
1	0	1	1	0	1	0	1	0	0
1	1	0	1	1	0	0	0	1	1

Compatibility boxes entirely filled with 0 will produce constrains 4 incompatible with system 1, i.e. the aggregated system 1 and 4 will have no solutions. Thus, the resolution for the given SubGI instance is “NO”.

**Cycle vs path:** Let input and pattern be the following digraphs:

$$G : 1 \rightarrow 2 \rightarrow 3 \rightarrow 1, S : 1 \rightarrow 2 \rightarrow 3$$

Compatibility matrix  $C$  for this SubGI instance looks as follows:

<i>1</i>	0	0	0	0	<i>1</i>	0	0	1	<i>1</i>
0	1	0	0	0	0	1	1	0	1
0	0	<b>1</b>	<b>1</b>	0	0	0	1	<b>1</b>	0
0	0	<b>1</b>	<b>1</b>	0	0	0	0	<b>1</b>	0
<i>1</i>	0	0	0	0	<i>1</i>	0	0	0	<i>1</i>
0	1	0	0	0	0	1	1	0	0
0	1	1	0	0	0	1	1	0	0
1	0	<b>1</b>	<b>1</b>	0	0	0	0	<b>1</b>	0
<i>1</i>	1	0	0	0	<i>1</i>	0	0	0	<i>1</i>

Constrains 4 produced by this compatibility matrix are compatible with system 1, i.e. the aggregated system has solutions. Two of the three solution grids of the system are shown in the above matrix in italic and in bold. Thus, the resolution for the given SubGI instance is “YES”.

**Cycle vs cycle:** Let input and pattern be the following digraphs:

$$G : 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1, S : 1 \rightarrow 2 \rightarrow 3 \rightarrow 1$$

Compatibility matrix  $C$  for this SubGI instance

looks as follows:

1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	1
0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	1
0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1	0	1
0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	1	1	0
0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	1
1	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1
0	1	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1
0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0
0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1
0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	0
1	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0
1	0	1	1	1	0	1	1	1	1	0	1	1	0	1	0	1	0
1	1	0	1	1	1	0	1	1	1	0	1	0	0	1	0	1	0
1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	0	1	1

Constrains 4 produced by this compatibility matrix are incompatible with system 1, i.e. the aggregated system has no solutions. To see that, let's apply system 1 to the compatibility matrix as constrains on its elements. To satisfy these constrains at least partially, the forth box column and the forth box row of the compatibility matrix have to be trimmed/depleted as follows:

1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
0	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0
0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0
0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0
1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1
0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0
0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0
1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0
0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1

After this depletion, the fact that the fourth box column contradicts with the third group of equations in system 1 becomes obvious. Thus, the resolution for the given SubGI instance is "NO".

## 6 Linear program for TSP

Let input  $G$  be an arc-weighted digraph, i.e. let each arc in  $G$  have a weight. TSP is a problem of finding a Hamiltonian cycle in  $G$  with the minimal total weight<sup>3</sup>. That is a NP-complete problem [2].

<sup>3</sup>Because  $G$  is a digraph, we actually consider here the Asymmetric Traveling Salesman Problem (ATSP).

The pattern  $S$  for TSP is any circular permutation matrix, for example:

$$S = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{pmatrix}_{n \times n}$$

Let's construct system 4 for SubGI instance  $(G, S)$ . Then, aggregated linear system 1 and 4 will express the Hamiltonian Cycle Problem which is a NP-complete problem [2], as well.

Let  $w(\mu, \nu)$  be a weight function - the weight of the arc from vertex  $\mu$  into vertex  $\nu$  in input  $G$ . As usual, let  $w(\mu, \nu) = +\infty$  for non-adjacent vertices. Then, the following asymmetric polynomial size linear program will express TSP:

$$\sum_{i,j,\mu,\nu} w(\mu, \nu) x_{ij\mu\nu} \rightarrow \min$$

- subject to constrains 1 and 4.

From the practical perspective, let's notice that we do not require function  $w(\mu, \nu)$  to be positive.

## 7 Linear model for SAT

In 1971, Cook [3] found that with a polynomial number of operations any non-deterministic Turing machine (NDTM) can be expressed by the appropriate conjunctive normal form (CNF): the question of whether there is an acceptable input is a question of whether the appropriate CNF is satisfiable. That made SAT the first NP-complete problem, because it is a NP-problem and the very words "NP-problem" mean a problem which can be solved by NDTM in polynomial time. In 1973, Levin [11] independently repeated the result in terms of search. In 1972, Karp [2] selected SAT as a root of NP-completeness theory: a problem is NP-complete if SAT can be reduced to that problem in polynomial time, and visa versa.

Let  $f$  be a given CNF:

$$f = c_1 \wedge c_2 \wedge \dots \wedge c_m$$

- where clause  $c_i$  is a disjunction of  $k_i$  literals - some Boolean variables or their negations. Formula  $f$  defines an instance of SAT: whether there is such a true-assignment to the involved Boolean variables which would make  $f = \text{true}$ .

Ultimately, we could apply the distributive laws and rewrite formula  $f$  in a disjunctive form (DF). That would reduce SAT to an existence problem for implicants in the emerging DF. This last problem can be easily expressed as a SubGI instance.

Let's enumerate literals in each of the clauses in formula  $f$ . For each couple of clauses  $(c_i, c_j)$ , let's build a compatibility box: the  $(\alpha, \beta)$ -element in the

matrix is 0 or 1 depending on whether the  $\alpha$ -th literal in clause  $c_i$  and the  $\beta$ -th literal in clause  $c_j$  are complementary. Let's aggregate all these compatibility boxes in a box matrix. Obviously, there is an implicant in the DF of  $f$  iff there is a grid of elements in the box matrix, one element per compatibility box, whose all elements are equal 1. Each such grid of elements consists of the couples of literals which participate in an implicant.

The box matrix built in such a way may be seen as input  $G$ . Then, pattern  $S$  may be a box matrix of the same structure as  $G$  but whose boxes are entirely filled with 0 except their upper-left-corner elements, which are equal 1:

$$S = (S_{ij})_{m \times m}, S_{ij} = \begin{pmatrix} 1 & 0 & \dots \\ 0 & 0 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}_{k_i \times k_j}$$

There is one obvious restriction on the relabeling of  $S$ : the elements of boxes  $S_{ij}$  are not allowed to leave their boxes. This restriction can be accommodated in system 4 with a polynomial number of additional linear constraints.

## Conclusion

We described a polynomial time reduction of SubGI to a polynomial size asymmetric linear system. The system consists of systems 1 and 4. Subsystem 1 depends on the size of SubGI instance, only. Subsystem 4 describes the structure of the given input and pattern. The system's asymmetry is due to the explicit involvement of the input and pattern's adjacency matrices in the construction of system 4 - see definition 3. So, the result may be seen as complementary to the Yannakakis theorem [1].

Linear system 1, 4 defines a sub-polytope in the Birkhoff polytope. Vertices of this sub-polytope are those permutation matrices which satisfy quadratic integral system 2. Relabeling of the input and pattern rotates this sub-polytope all over the Birkhoff polytope.

Ultimately, system 1, 4 may be seen as a parallel testing of all guesses, where guesses are  $n \times n$  permutation matrices - the unknowns in system 2. Basically, this parallelization was achieved with encoding SubGI in the contradictions between relabeling possibilities for different vertices.

Obviously, the described "continuous" solution of SubGI is not unique. Also, we could develop a polynomial time discrete algorithm which would search the compatibility matrix for the solution grids as, for example, it was done in [12] for 3SAT which is a NP-complete problem [3], too<sup>4</sup>.

## References

- [1] Mihalis Yannakakis, *Expressing combinatorial optimization problems by linear programs*, In Proc. of the twentieth annual ACM Sympos. on Theory of computing, Chicago, Illinois, pp. 223 - 228, 1988
- [2] Richard M. Karp, *Reducibility Among Combinatorial Problems*, In Complexity of Computer Computations, Proc. Sympos. IBM, Thomas J. Watson Res. Center, Yorktown Heights, N.Y. New York: Plenum, pp. 85 - 103, 1972
- [3] Stephen Cook, *The complexity of theorem-proving procedures*, In Conference Record of Third Annual ACM Symposium on Theory of Computing, p.151-158, 1971
- [4] J.R. Ullmann, *An algorithm for subgraph isomorphism*, Journal of the ACM, 23(1):31-42, 1976
- [5] J.E. Hopcroft and J.K. Wong, *Linear time algorithm for isomorphism of planar graphs*, In Annual ACM Symposium on Theory of Computing, p.172 - 184, 1974
- [6] E.M. Luks, *Isomorphism of graphs of bounded valence can be tested in polynomial time*, Journal of Computer and System Sciences, p. 42 - 65, 1982
- [7] D. Eppstein, *Subgraph Isomorphism in Planar Graphs and Related Problems*, 6th ACM - SIAM Symp. Discrete Algorithms, San Francisco, 1995, p. 632 - 640.
- [8] L. G. Khachiyan, *A polynomial algorithm in linear programming*, Doklady Akademii Nauk SSSR, 224 (1979), pp. 1093-1096.
- [9] M. Grotschel, L. Lovasz, A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer, Berlin, 1988
- [10] G. Birkhoff, *Tres observaciones sobre el algebra lineal*, Univ. Nac. Tucumán Rev, Ser. A, no. 5, (1946) p. 147-151
- [11] Levin, Leonid, (1973). *Universal search problems (Russian: Universal'nye perebornye zadachi)*, Problems of Information Transmission (Russian; Problemy Peredachi Informatsii) 9 (3): 265-266, 1973 (Russian); translated into English by Trakhtenbrot, B. A. (1984). *A survey of Russian approaches to perebor (brute-force searches) algorithms*, Annals of the History of Computing 6 (4): 384-400. doi:10.1109/MAHC.1984.10036.
- [12] Sergey Gubin, *Polynomial time algorithm for 3-SAT*, arXiv:cs/0701023 [cs.CC]

<sup>4</sup>For 3SAT, see a demo at <http://www.timescube.com>